

VU Research Portal

On blind tests and spatial prediction models

Fabbri, A.G.; Chung, C.F.

published in

Natural Resources Research
2008

DOI (link to publisher)

[10.1007/s11053-008-9072-y](https://doi.org/10.1007/s11053-008-9072-y)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Fabbri, A. G., & Chung, C. F. (2008). On blind tests and spatial prediction models. *Natural Resources Research*, 17(2), 107-118. <https://doi.org/10.1007/s11053-008-9072-y>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

On Blind Tests and Spatial Prediction Models

Andrea G. Fabbri^{1,2,4} and Chang-Jo Chung³

Received 10 December 2007; accepted 30 April 2008
Published online: 28 May 2008

This contribution discusses the usage of blind tests, BT, to cross-validate and interpret the results of predictions by statistical models applied to spatial databases. Models such as Bayesian probability, empirical likelihood ratio, fuzzy sets, or neural networks were and are being applied to identify areas likely to contain events such as undiscovered mineral resources, zones of high natural hazard, or sites with high potential environmental impact. By processing the information in a spatial database, the models establish the relationships between the distribution of known events and their contextual settings, described by both thematic and continuous data layers. The relationships are to locate situations where similar events are likely to occur. Maps of predicted relative resource potential or of relative hazard/impact levels are generated. They consist of relative values that need careful quantitative scrutiny to be interpreted for taking decisions on further action in exploration or on hazard/impact mitigation and avoidance. The only meaning of such relative values is their rank. Obviously, to assess the reliability of the predicted ranks, tests are indispensable. This is also a consequence of the impracticality of waiting for the future to reveal the goodness of our prediction. During the past decade only a few attempts have been made by some researchers to cross-validate the results of spatial predictions. Furthermore, assumptions and applications of cross-validations differ considerably in a number of recent case studies. A perspective for all such experiments is provided using two specific examples, one in mineral exploration and the other in landslide hazard, to answer the fundamental question: how good is my prediction?

KEY WORDS: Blind test, cross-validation, spatial prediction, favorability, mineral exploration, hazard.

INTRODUCTION

Constructing a map is an exercise to capture the distribution of observed objects or measurements, in conveniently simplified terms, to comprehend and communicate their spatial significance. In the geosciences such objects are considered as indicators of processes of concern, either physical or social, or

combined. During the past half a century, mapped objects have become more factual, i.e., less interpretative, they comprise increasingly less manually-captured information with the development of remote sensors, and their representation has turned from analog to digital. The latter is a fact that has encouraged the practice of overlaying different types of maps covering a same study area to derive specific themes with combined features of varying spatial continuity, connectivity, and other desirable spatial properties.

For example, the spatial setting of an observed distribution of mineral occurrences could be recognised as a “non-random” distribution over preferred combinations of mapping units, thereby partly revealing their environment of deposition. This was

¹Università di Milano-Bicocca, Piazza della Scienza, 1, 20126, Milan, Italy.

²Vrije Universiteit, de Boelelaan 1087, 1081 HV, Amsterdam, The Netherlands.

³Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8.

⁴To whom correspondence should be addressed; e-mail: andrea.fabbri@unimib.it

the starting point of many applications of statistical models to predict future discoveries in mineral exploration.

Some of the drawbacks of such experiments, however, have been that: (1) the established relationships are limited to the study area selected and its assumed time relevance, thus providing only relative characterization; (2) several quantitative models and associated assumptions can be used to express the spatial relationships in different ways; and (3) the quality of the prediction results is difficult or sometime impossible to assess.

Similar considerations can be made for the spatial prediction of natural hazards and of environmental impacts. There, too, map data layers of thematic units and continuous values are overlaid and the resulting aggregated values are transformed and modeled to express the likelihood of hazard or of impact occurrence, so that study areas can indicate priority locations for detailed inspection in view of hazard prevention, avoidance, or mitigation.

This contribution discusses how empirical measures of relative quality, termed cross-validation, can be and should be obtained through blind tests, BT, of spatial predicted values from map overlays. Such measures require specific assumptions, scenarios, and analytical strategies. For simplicity we will use the term *event occurrence* to refer equally to resource discovery or hazard-impact, even if the former implies a process already occurred (deposition) while the latter a process to occur in the future.

SPATIAL PREDICTION MODELS AND ASSOCIATED ASSUMPTIONS

Various statistical models can be used to establish spatial relationships between the distribution of point-like or patch-like events, and the mapping units in which they tend to occur. The latter can be categorical, such as lithologies or land uses, or can express continuous values, such as geophysical anomalies or terrain slope values. The events are preferably instances of a specific type so that consistency of origin and context can be expected when relating them to the categorical units and continuous value maps.

Commonly used spatial prediction models are based on: (1) Bayesian Probability Theory's Joint Conditional Probability function and the Likelihood Ratio function, and its derived monotonic functions such as the Certainty Factor, and the Weights of Evidence; (2) Zadeh's Fuzzy Set membership

function; and (3) Dempster-Shafer Evidential Theory's Belief Function. A mathematical unified framework for these models has been provided by Chung and Fabbri (1993), together with criteria to construct them and to estimate predicted values. One main assumption of the above spatial prediction models is that each map data layer provides "independent" evidence of favorable setting. A general term used for the models is Favorability Functions.

An additional assumption that support the application of the models to predict further discoveries or future hazards is a degree of similarity between the observed-constructed settings of the known event occurrences and those of the future ones. It is that "degree of similarity" that allows extrapolation in space and possibly in time.

Another set of inherent assumptions is the causal relationships between the mapping units and the events. Such relationships are the result of expert knowledge, i.e., of the opinion of scientists specialized on the commodities or the hazards. The experts are to provide guidance for the construction of the spatial databases and for their interpretation. Another general assumption is that the spatial database constructed for a study area sufficiently documents the above relationships so that the statistics obtained from it can be used to support the spatial prediction. Inherent to this assumption is a degree of uniformity of detail and consistency or "granularity" between the map data layers. Such layers in general consist of a mixture of categorical and continuous values.

RELATIVE INDEXES AND THEIR MEASURES

The statistics from the spatial database are considered as partial evidence in favor or against the occurrence of events. The assumptions on the relevance of those statistics to represent the condition of future occurrences provide essentially a way to obtain a relative ranking of all units within a map and later of all points with aggregate values ranked from a set of overlaid map layers, based on a spatial prediction model. Using the model means that, given two separate points in a study area, one point will have a higher aggregate value than the other. The relative ranks are the only interpretable evidence obtainable from the model and the database. It is doubtful that the original scores have any direct meaning other than the relative ranking.

After constructing a favorability function as the spatial model, a relative potential level is estimated at every pixel by computing the score of the favorability function at every pixel in the study area. We will be using the term “potential” to refer to either “resource discovery” or to “hazard” to indicate the relative predicted index scores. These computed scores normally range from 0 to infinity. Because they express relative levels of potential, they can be replaced by ranks (or orders) instead of the actual scores. In a study area, suppose that there are n pixels. We expect to have n estimated scores, one at each pixel. These n values are sorted in decreasing order and replaced by their rank, ranging now from 1 to n . Dividing each rank by the number of pixels n standardizes the ranks. The resulting standardized ranks range from $1/n$ to 1 were termed as “predicted relative potential indices,” or PRP indices, with the pixel having the highest PRP index being assigned the value 1, and the pixel having the lowest PRP index being assigned a value of $1/n$. By plotting the PRP index at each pixel, a PRP map is generated. To illustrate the PRP index, let us consider a pixel with 0.95 as the index. It means that the pixels whose favorability function scores are greater than the score of the pixel with 0.95 as the PRP index cover 5% of the study area. We will later use the PRP indices to evaluate the prediction maps through “fitting-rate curves” and “prediction-rate curves” using cross-validation.

Simple ways to analyze rank statistics were discussed by Chung and Fabbri (2003) who have described several benefits of using such a ranking procedure to generate the potential classes for a prediction map. For example, suppose that we wish to generate 100 equal-size prediction classes where each class covers 1% of the study area. Then such PRP indices obtained by the ranking procedure provide a useful tool. The 100 equal-size classes are generated in the following manner. Assign “Class 100” consisting of the pixels with the PRP indices larger than 0.99 and less than or equal to 1. The pixels in “Class 100” cover 1% of the study area. “Class 99” is assigned the pixels with the predicted potential indexes larger than 0.98 and less than or equal to 0.99. Similarly “Class 1” consists of the pixels with the PRP indices less than or equal to 0.01. When considering an appropriate number of potential classes, however, the meaningful number of classes depends on the quality of information available in the database and on the significance of the model used. To illustrate the relationship

between computed favorability function values and corresponding PRP indices, a scatter plot can be used.

The first step to evaluate a prediction map is to compare the predicted potential indices of the pixels with the known occurrences of the events (note that these events were used to generate the prediction map) and such comparison generates the “fitting-rate curve” of the prediction map. Suppose that we have m known events. To produce the fitting rate curve of a prediction map, simply obtain m predicted potential indexes at m known events and then sort m values in decreasing order; (q_1, q_2, \dots, q_m) , where q_1 indicates the largest PRP index. We generate the following m pairs:

$$\{(1 - q_1), 1/m\}, \{(1 - q_2), 2/m\}, \dots, \{(1 - q_m), 1\}$$

The scatter plot of these m pairs constitutes the fitting-rate curve where the X -axis represents the portion of the study area assigned to a “potential” class and the Y -axis represents the proportion of the known events that have occurred within the assigned “potential” class. Such a fitting, however, only reflects how the classes discriminate between the settings identified using the distribution of the observed events, and does not necessarily reflect the distribution of future occurrences. For that, other techniques and assumptions are necessary, as we will see later on through the blind-test procedures.

HOW GOOD IS THE PREDICTED RELATIVE POTENTIAL INDEX AS A PREDICTOR?

Potential indices as we have described, are to reflect not just the fitting to the prediction classes but the likelihood of future event occurrence, given the combined presence of the map unit data layers. Such a likelihood, however, is restricted so far to the distribution of the past events and the associated database of the study area. To study and interpret their effectiveness as predictors of future occurrences we have to assume a similarity of conditions between what has been observed in the past and what will occur in the future (e.g., new resource discoveries, new hazardous events, etc.). Saying that the past is the key to the future is only a starting point that means that we are willing to infer, given the observed trends, that within a given time interval and within a given study area, we expect as many events (or say twice as many, or some other larger or

smaller number) as observed in the database. Alternatively but impractically, we could wait for a sufficiently long time and see how many events would occur with respect to our prediction.

Another more convenient empirical way to study the effectiveness of our initial prediction that used the distribution of all past events is to perform a cross-validation of the prediction results by partitioning the set of observed events into a prediction subset and a testing subset. With the former we can obtain a second prediction and the relative ranked equal area classes. With the latter we can verify how the testing subset of events is distributed across those new classes. A “good” prediction should show a strong clustering of testing events in the higher value classes. This second clustering, however, will be different from that of the fitting classes mentioned earlier. Nevertheless, it is a measure of its effectiveness.

The next section describes how to interpret the prediction results via blind tests.

WHAT IS A BLIND TEST AND WHAT IS IT TELLING US?

A BT is a fundamental way to cross-validate the results of spatial predictions empirically, short of waiting for events to occur. A BT is obtained, for instance, by pretending that part of the known events is unknown. It will be used to test the prediction results generated using the other part of the known events to establish the spatial relationships. The probability table estimated via BT depends entirely on how the partition is selected and the interpretation of the probability is again solely contingent and subject to the partition. The event partition can be obtained in various ways, depending on the quality and quantity of the event data available.

(i) Only Very Few Events are Known that Cannot be Separated in Different Periods or Sub-areas

One event out of m is used to BT and all the $m - 1$ remaining ones are to generate a prediction to be cross-validated by the excluded event. Using the $m - 1$ remaining events, a prediction map based on the PRP indices is constructed. The PRP index is obtained at the pixel containing the excluded event. The operation is iterated m times, once for each of the m excluded events. This leads to generating m

PRP indices showing how well each future event can be predicted, as the “next” event to occur, by all the other existing ones. To produce the “prediction-rate curve,” simply sort the m indices in decreasing order; (p_1, p_2, \dots, p_m) , where p_1 indicates the largest PRP index. We generate the following m pairs:

$$\{(1 - p_1), 1/m\}, \{(1 - p_2), 2/m\}, \dots, \{(1 - p_m), 1\}$$

The scatter plot of these m pairs constitutes the prediction rate curve where the X -axis represents the proportion of the study area assigned to a “potential” class and the Y -axis may be regarded as the representation of the proportion of the “future” events that have occurred within the assigned “potential” class. In contrast with the fitting-rate curve that only reflects how the classes discriminate between the settings identified using the distribution of the observed events, the prediction-rate curve reflects the distribution of future occurrences in the prediction map.

(ii) Numerous Events are Known but Cannot be Separated in Different Periods or Sub-areas

A random half of the events is used to BT and the other random half is used to predict. The BT can be repeated inverting the role of the two random halves or it can be repeated several times with newly generated random halves, to obtain integrated statistics on the stability and reliability of the prediction results.

(iii) Numerous Events are Known that can be Separated in Several Temporal Sub-groups

A BT is performed using the older set of events to predict and the younger set for testing. The statistics from the BTs provide true time prediction results. In such cases the quality of the prediction results should reflect the stability in time of the thematic map units subjected to transformation (e.g., climatic or human-induced) such as land use or land cover.

(iv) Numerous Events are Known that can be Separated in Several Spatial Sub-groups

The event distribution in some sub-areas is used to BT the results of a prediction obtained from an

adjacent sub-area, in which the spatial relationships have been established. It means that the statistics on the relationships is obtained from one area and then is applied to another area. The BT is dependent on the similarity of conditions and events in the areas analyzed and compared. In some situations, for instance, the spatial data allow a combination of Strategies (iii) and (iv).

(v) Other Types of BTs can be Performed

Changing the combination of thematic and continuous data layers or the quality-resolution, BT are obtained in experiments corresponding to one or more of the previous types of BTs just described.

To produce the “prediction-rate curve” for (ii), (iii), (iv) and (v), as described in (i), we have to obtain PRP indices from the pixels that contain the observed events but that were not used in constructing the prediction map in BT. Supposing that we obtain k indices and sort them in decreasing order; (p_1, p_2, \dots, p_k) , where p_1 indicates the largest predicted potential index. We generate the following k pairs:

$$\{(1 - p_1), 1/k\}, \{(1 - p_2), 2/k\}, \dots, \{(1 - p_k), 1\}$$

The scatter plot of these k pairs constitutes the prediction rate-curve where the X -axis represents the proportion of the study area assigned to a “potential” class and the Y -axis may be regarded as the representation for the proportion of the “future” events, which occurred within the “potential” class. Performing BTs appears as a practical way of interpreting many aspects of prediction modeling: (1) quality of data layers (categorical and continuous), distribution of types of known events/discoveries, and expert’s knowledge of the spatial database; (2) significance of the predicted relative PRP index maps; (3) effect of database partitioning in modeling; (4) comparisons of the results of different prediction models; and (5) assessment of scenarios for exploration or for risk analysis.

A general purpose strategy for favorability function predictive modeling is shown in Fig. 1 as an operational flowchart with three stages. The distribution of known discoveries or of the hazardous occurrences is used to establish their spatial relationship with the units of the input map data layers. The terms discoveries or occurrences are used interchangeably to refer to exploration or to hazard/impact applications. The interpretation of the

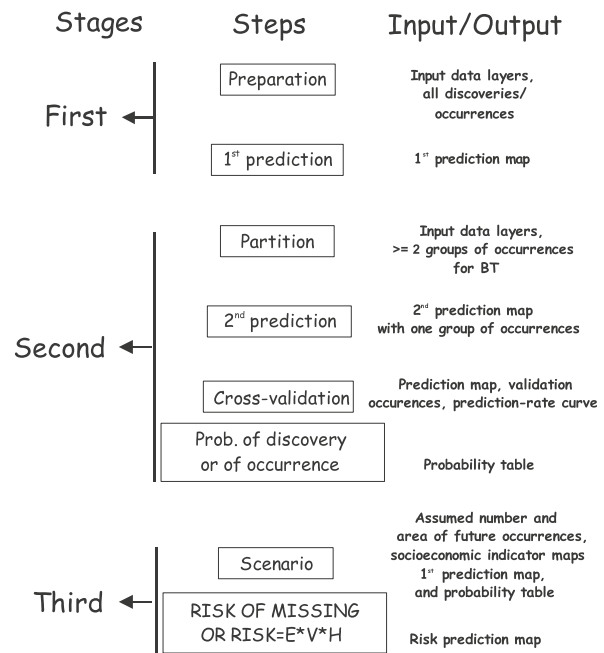


Figure 1. The three stages of favorability function modeling. The probability table from the Second Stage, which depends entirely on how the partition for BT is made for validation, is the most critical piece of information to interpret the prediction results in the First Stage and to obtain the risk prediction map in the Third Stage, where E is element at risk, V is vulnerability, and H is hazard. The term discovery is used interchangeably with the term occurrence to refer to exploration and to hazard/impact applications, respectively.

probability table obtained depends entirely on how the partition for BT was made. To perform analyses according to the strategies listed earlier, iterations can be executed looping back one or more steps. In the next section examples of applications with and without BTs are discussed. Dedicated software based on cross-validation has been discussed by Fabbri, Chung, and Jang (2004).

SPATIAL PREDICTIONS WITH EVENT PARTITIONS AND BLIND TESTS

Some General Purpose Applications

Once a unified framework for favorability function models had been set up (Chung and Fabbri, 1993) and applications of various models were developed, it became evident that to interpret the results of predictions, either mineral potential maps or hazard maps, empirical tests were necessary to obtain scientific measures of success and decision

values of the prediction results. BTs were used, for instance, in cross-validations for the following purposes:

- assessment of predictive power of landslide hazard (Chung, Fabbri, and van Westen, 1995), a first application of BT to interpret the “goodness” of spatial prediction results;
- comparisons of the performance of different prediction models, and their integration with expert’s knowledge (Chung and Fabbri, 1998, 1999);
- estimation of probability of mineral discovery by an operational unit area for exploration (Chung, Fabbri, and Chi, 2002a);
- separation of influential and non-influential data layers in landslide hazard predictions (Chung, Kojima, Fabbri, 2002b); it enabled to identify predictions of greater reliability due to the higher empirical support to characterize the settings of landslide occurrence;
- assessment of uncertainty in landslide hazard predictions (Chung and others, 2006); by iterating many times the selection of random halves of the events, prediction-rates were obtained to express the level of uncertainty associated with the predicted classes;
- comparisons between spatial, temporal, and spatial/temporal predictions (Chung and Fabbri, 2008);
- cost-benefit analysis of prediction-rate curves of landslide hazard (Chung and Fabbri, 2003); a ratio of effectiveness was applied to identify the most reliable parts of the prediction-rate curves;
- landslide risk assessment via probability of occurrence estimation (Chung and Fabbri, 2004; Chung and others, 2005a); the introduction of socioeconomic indicator maps led to the assessment of landslide risks to people, infrastructures, and valuable land uses.

Two Examples of BT Strategies

To clarify in some detail the usefulness of BT, one recent application of spatial prediction modeling in mineral exploration, with only six known discoveries, is discussed, followed by a second application to landslide hazard for which 92 known occurrences are used. The two BT strategies are

different, so are the results obtained and their significance.

A spatial database for diamond exploration in the Lac de Gras area of the Northwest Territories, in Canada, was used by Chung and others (2005b) and Chung and Fabbri (2005) to obtain the prediction-rate curves shown in Fig. 2. The study area covers 34.6×22.9 km (692×450 pixels of 50 m resolution) and contains six diamondiferous kimberlite ore bodies (Beartooth, Panda, Koala, Koala North, Fox and Misery). Additionally, 15 kimberlites with only micro-diamonds were known. Radiometric and magnetic sensor maps interpolated from parallel flights, proximity maps to faults and dikes (as continuous data layers) and the presence of two indicator minerals, chromium-spinel and chromium-diopside, were used in the study. In addition, a bedrock lithology map (categorical data layer) was employed to characterize the spatial associations of the ore bodies and of the other kimberlites with micro-diamonds.

A fuzzy set prediction model based on the likelihood ratio function was instrumental to obtain and interpret the prediction maps following strategy (i) in section “What is a Blind Test and What is it Telling Us?”. A first prediction map was obtained using the locations of all the six deposits. It was then interpreted with the prediction table estimated from the cross-validation procedures using six blind tests. Six more prediction maps were so obtained from the BT. Figure 2 shows parts of the prediction-rate curve in blue from the latter six prediction maps. For a comparison, two additional experiments with different inputs were performed: (1) instead of seven data layers, only one data layer, the magnetic total field, with the six ore bodies was used in an additional BT, using the same strategy (i) in section “What is a Blind Test and What is it Telling Us?”, to study the effects of input data layers, and (2) the same seven data layers in the earlier BT, but using the 15 kimberlites with micro-diamonds instead of the six ore bodies, to test whether kimberlites with micro-diamonds can “predict” the locations of the six ore bodies. The cross-validation results were also plotted in Fig. 2.

As discussed in Chung and others (2005b), even without seeing the 13 prediction maps generated, we can compare in Fig. 2 the prediction results. The comparison is made by considering the area proportion of the higher prediction classes containing the ore bodies, each predicted as “next” to occur by the other five, as the blue and the red prediction-rate

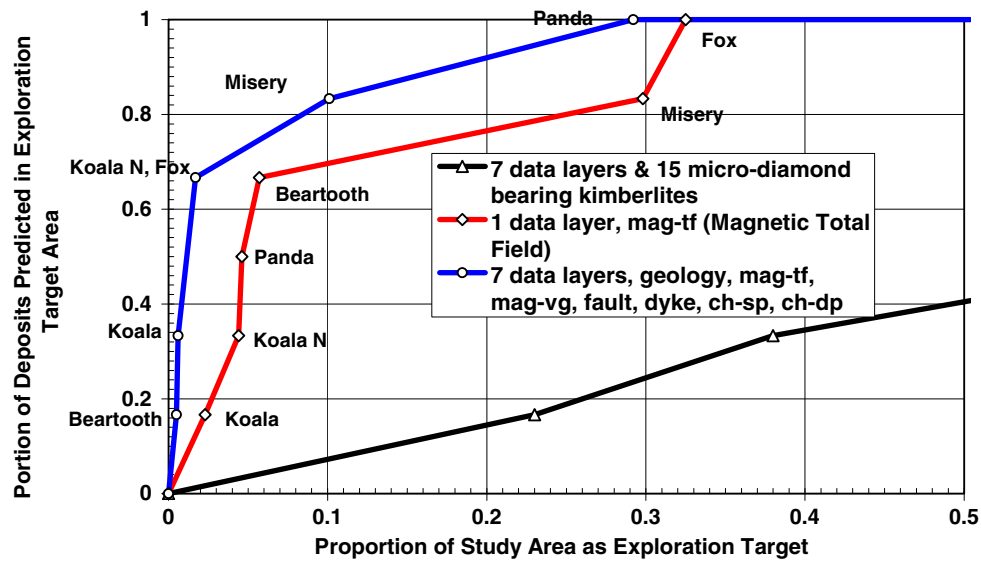


Figure 2. Example of prediction-rate curves obtained from cross-validation in the Lac de Grass area, Northwest Territories, Canada, obtained using strategy (i) in section “What is a Blind Test and What is it Telling Us?”. The cumulative plots allow the probability of discovery of a new deposit location to be computed within the high potential area of the corresponding prediction maps, not shown here (Chung and others, 2005b). Vertical gradient is vg, total field is tf, chromite is ch, spinel is sp, and diopside is dp.

curves. Obviously, the prediction of the six ore bodies by the locations of the 15 kimberlites with micro-diamonds is poor! The BT shows that statistically the two types of kimberlites have different characterizations. It suggests that the locations of kimberlites with micro-diamonds do not provide any useful information to locate undiscovered ore bodies in this case study. In a second application, to hazard modeling, a greater number of known occurrences allowed a different strategy to be selected.

A spatial database for landslide hazard studies was constructed for the Fanhões-Trancão area, north of Lisbon, in Portugal. The study area is 13.3 km². Detailed geologic-geomorphologic mapping at 1:2,000 identified 92 shallow translational slides. They were compiled and digitized into a 5 × 5 m resolution spatial database consisting of digital images of 760 × 700 pixels. The causal factors (i.e., related to the occurrence of landslides) are: continuous data layers, i.e., elevation, slope angle, aspect angle obtained from the digital elevation model (DEM), and categorical ones, i.e., geology map, surficial deposit map, and land-use map.

The 92 landslides in the study area consist of 43 pre-1980 landslides and of the remaining 49 post-1980 landslides. The region has been the focus of numerous geomorphologic analyses for hazard zonation by Zêzere and others (2004).

A landslide hazard (potential) prediction map of the Fanhões-Trancão area, Portugal, was obtained by Chung and Fabbri (2008), using the Fuzzy Set membership function of the Likelihood Ratio Function. The same function has been used in the other prediction experiments. Input data were the locations of the polygons of the 92 shallow translational landslides and the six geomorphologic and topographic map layers. In that application, the number of the 92 landslide polygons that fell into each of 200 hazard classes was counted. Each class corresponded to 0.5% of the study area. To be counted within a class, at least 50% of the pixels in a landslide polygon must be included in the class. The counts are weighted by the numbers of pixels in the polygons. Weighted counts of the landslide polygons form the “fitting-rate table” or curve that was plotted as the gray line with triangles in Fig. 3 with the horizontal axis representing the proportion of the study area predicted as hazardous, and the vertical axis showing the cumulative proportion of landslides falling within each class. A second experiment generated another prediction map using only 43 pre-1979 landslides and its fitting-rate curve is also shown in Fig. 3, falling below the previous fitting-rate curve based on all the 92 landslides. The third curve in the illustration is the prediction-rate curve from the latter experiment that provides a measure

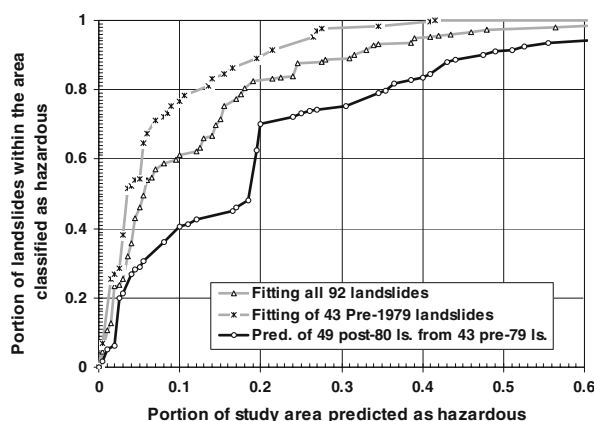


Figure 3. Fitting-rate and prediction-rate curves of landslide hazard prediction maps in the Fanhões-Trancão area, Lisbon, Portugal (modified after Chung and Fabbri, 2008).

of “goodness” of the classes obtained in the two preceding predictions using the time partition of the landslide occurrences. Strategy (iii) of section “What is a Blind Test and What is it Telling Us?” was used in this experiment. Here the assumption was made that the 49 post-1980 landslides are unknown and represent the future occurrences during a 25-year period (1980–2004). Additionally, we assumed that the prediction rate obtained represents the prediction power of the first prediction that used all the 92 occurrences for the next 25 years, i.e., the period 2005–2030.

The 10% of the study area with the highest predicted values (Fig. 3) corresponds to a prediction rate of 41% whereas the fitting rates are 61 and 77%, respectively. The latter two would overestimate the “goodness” of the prediction. They only indicate the “goodness” of fit between the landslides and the causal factors.

In another experiment, the study area was divided into two mutually exclusive sub-areas, the left region and the right region, as in strategy (iv) of section “What is a Blind Test and What is it Telling Us?”. An experiment of this type would enable the similarity of geomorphologic settings or of climatic conditions to be tested. The left region contains 38 landslides (13 pre-1979 and 25 post-1980) and the right region includes 54 landslides (30 pre-1979 and 24 post-1980). Lower prediction-rate curves are compared (Fig. 4) to the prediction-rate curve from Fig. 3. There the previous prediction of the 49 post-1980 from the 43 pre-1979 landslides is compared with two spatial predictions in the right half using the landslides from the left half region and vice

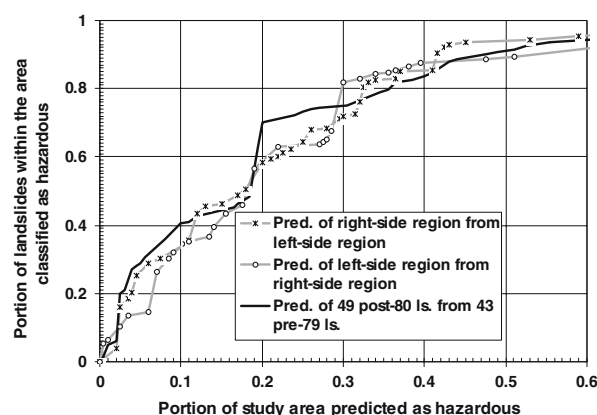


Figure 4. Some prediction-rate curves of landslide hazard prediction maps in the Fanhões-Trancão area, Lisbon, Portugal (modified after Chung and Fabbri, 2008).

versa. Corresponding values for the 10% of highest predicted classes are 41 vs. 37%. A mosaic of two prediction images is the result of this validated prediction image.

An extensive discussion of these and more experiments can be found in Chung and Fabbri (2008), who also combined strategies (iii) and (iv) that provided even lower prediction-rate values.

Clearly, all the above-mentioned characteristics of “goodness” of the prediction images generated would be totally unknown without cross-validation via BT. Consequently, the BTs lead to considerations and introspections on the similarity of occurrences in time, of settings in time and in space, of comparability between adjacent study areas, between prediction models, and on how to use the prediction-rate values for estimating the probabilities of occurrences for each class or for each pixel. Far from trivial consequences follow the use of BT!

CONSIDERATIONS ON RECENT SPATIAL PREDICTIONS IN THE GEOSCIENCES

Having explored the information that must be extracted from spatial databases by BT of the prediction results, it should be instructive to consider a number of research papers in spatial modeling that would greatly benefit from BT and/or from more extensive applications of BT. Since cross-validations of spatial prediction results have been initially proposed (Chung, Fabbri, and van Westen, 1995; Chung and Fabbri, 1999), relatively few examples of BT can

be found either in mineral exploration or in natural hazard studies.

In the past 12 years or so interest in empirical validation or BT for prediction modeling in mineral exploration has varied from complete absence to considerable concern. However, there does not seem to be a consistent systematic or standardized approach to the application of cross-validation techniques. For instance, the evaluation of spatial modeling for epithermal gold deposit prediction by Raines (1999) rightly saw the prediction results as a “relative ordinal rank” but no BT was reported. The separation of favorability values into favorable, permissive, and non-permissive was obtained by identifying breaks in the cumulative area ranks. That corresponds to using the fitting rates of the deposits used to predict and not to the prediction-rates from a cross-validation.

A different approach is the one by Singer and Kouba (1999) who compared several probabilistic models in the prediction of mineral deposits. They analyzed a test data set of 15 volcanic hosted massive sulfide deposits in a study area with 23 binary map data layers in the Province of Manitoba, Canada. Considered as wise by the authors was to perform independent validation tests by dividing the entire study area in two parts, one for predictive modeling and the other for validation. A random subset of 8 of the 15 deposits was selected together with a randomized half of the 6460 unique-condition polygons covering the study area and containing the 8 modeling deposits. The other half contained the remaining 7 deposits. Predictions were compared in terms of correctly classified polygons as deposit polygons or as barren polygons. Interestingly, they observed that very few deposits were correctly recognized in the independent tests whereas in the initial prediction modeling a high percentage had been recognized. Those authors made efforts to discuss in depth the pros and cons of the methods used, including the loss of information caused by binarizing all map data even when continuous. Nevertheless, also in that case, their analyses could be further expanded by applying strategy (i) of section “What is a Blind Test and What is it Telling Us?” (i.e., the take-one-out procedure) also used for the application described in Fig. 2.

An illustrative instance of a successful application is the one by Cheng (2004), who applied spatial modeling to predict the potential distribution of artesian aquifers in the Oak Ridge Moraine study area, near Toronto, Canada. As training points for

modeling he used the spatial distribution of 353 wells with water level above the surface. Binary expressions of surficial geology map, distance from thick drift layers, distance from the Oak Ridge Moraine, and distance from steep slope zones were used as evidential data layers. Buffer zones with unequal intervals were generated to obtain binary units from distance maps. The purpose was the identification of combinations of conditions to reduce the prediction areas of having flowing wells by two-thirds by generating a posterior probability map. BT of the results was not described, nevertheless the application would appear promising even if it cannot be certified how much so. It can be suggested that the use of strategy (ii) of section “What is a Blind Test and What is it Telling Us?” and the repetition of the analysis, say 30–40 times, with new random half partitions of the training and validation points would provide empirical means to interpret the “goodness” of the relative posterior probability value ranking obtained. In addition, a comparison of the 30–40 results would help to assess their robustness. The applications considered are just used here to exemplify the likely benefits of BT even in innovative and successful contributions, independently from the prediction models used.

Other more recent works dealt with problems such as the assessment of the quality of the prediction results (Porwal, Carranza, and Hale, 2003a, 2003b), and the comparison of different predicting methods and models when analyzing the same data set (de Quadros and others, 2006; Brown, Groves, and Gedeon, 2003; Porwal, Carranza, and Hale, 2006a, 2006b). Much of the emphasis in those works, however, was more on experimenting with new advanced techniques than on the interpretation of the significance of their application results. In addition, the strategies and specific assumptions of those cross-validations techniques were so different that it is not possible to evaluate or compare their usefulness in more general experiments or situations. For instance, lumping together fitting and prediction rates complicates the evaluation of the prediction quality. Thresholds to transform multi-value prediction maps into binary or tertiary maps are likely to weaken the cross-validation. In addition, some cross-validation experiments appeared limited to the training of classifiers and not directed to interpret the final prediction results.

Applications of spatial prediction models to natural hazard show a similar trend in the last few years. For instance in a special issue of *Natural*

Hazards there are contributions without validation of prediction results (e.g., Corominas and others, 2003), one in which validation has been avoided, in favor of fitting curves, with the argument of unavailability of the time of occurrence of landslides in the database (van Westen, Rengers, and Soeters, 2003), three studies in which validation was considered as integral part of the interpretation of hazard predictions (Santacana and others, 2003; Remondo and others, 2003a, 2003b), and two more studies in which validation was used to explore and compare prediction powers or to eliminate misunderstandings on perceived obstacles to spatial predictive modeling (Chung and Fabbri, 2003; Fabbri and others, 2003).

Indicative of the degree of confusion now remaining about validation in spatial prediction modeling is a recent collection of papers on spatial modeling in GIS. In this collection, four contributions deal with prediction of hazards (landslides) or vulnerability (aquifer), and six with the prediction of natural resources (metals, aggregates, and soils). All contributions claim to perform validations of modeling results; however, entirely different strategies are followed and assumptions made. Some approaches use fitting-rate curves (success rates) to identify “natural breaks” in them and obtain interpretable classes (Arthur and others, 2007; Masetti, Poli, and Sterlacchini, 2007; Poli and Sterlacchini, 2007; Behnia, 2007; Nelson, Connors, and Suárez, 2007). Generally weak comparisons are made of different prediction results by using either too few classes or too few occurrences to verify limited numbers of predictions (e.g., Nykanen and Ojala, 2007; Coolbaugh, Raines, and Zehner, 2007; Tissari and others, 2007). No effective validation of the prediction results appears in those contributions. Robinson and Larkin (2007) provide the only instance of prediction-rate curves in a diagram with proportions of sites correctly predicted (sensitivity) on the vertical axis and the cumulative area fraction (of study area) on the horizontal axis. Following a technique applied by Begueria (2006), they use a function of the area under the curve to establish the quality of the model prediction results. No further discussion is provided of the significance of such curve pattern in prediction modeling.

Applications that seem to lead to a more consistent approach of BT in mineral exploration are those by Chung (2003), Harris and others (2003), Agterberg and Bonham-Carter (2005), and Skabar (2005). Recent works on landslide hazard based on

cross-validations are the ones by Zêzere and others (2004) and by Lee and others (2006). In natural hazard studies the approach by Chung (2006) and Chung and Fabbri (2003, 2004, 2008) are targeting a more consistent way to use cross-validation techniques to estimate probabilities of occurrence of hazardous events.

CONCLUDING REMARKS

We have discussed how in spatial prediction modeling only relative ranks can be obtained using prediction models and their assumptions. We have dealt with the problem of assessing the “goodness” of the prediction results via a variety of empirical blind tests. A three-stage strategy for favorability function modeling has been proposed for which dedicated software is available that is soundly based on cross-validation. Examples of general purpose spatial prediction were listed, followed by two applications of BT that use prediction-rate curves to interpret the prediction results and proceed with the estimation of probabilities of occurrence. A number of recent applications were pointed out in which varying degrees and strategies of validation were attempted, while other ones seem to use ad hoc scenarios of limited effectiveness. Some additional applications appeared to potentially lead to a standardization of validation techniques.

A few recommendations can now be made for further research. In order to establish standards to interpret and compare the results of spatial predictions, three initiatives must be initiated in the geosciences: (1) identify one or two spatial databases to be distributed and analyzed by many researchers with different models to achieve agreement on how to construct BTs; (2) organize a special meeting on the standardization of validation strategies; and (3) focus on representing and assessing by BT the uncertainties associated with the prediction results. The authors of this contribution are committed to the last initiative.

There is now a wealth of different prediction methods and many applications have been attempted; however, scientific progress at present is perhaps needed more in assessing the significance and stability of the predictions obtained than in devising additional ways to establish spatial relationships with sophisticated new prediction models whose effectiveness may not be easily evaluated.

ACKNOWLEDGMENTS

The authors thank Prof. Carlos Roberto de Souza Filho, University of Campinas, Brazil, Dr. Graeme F. Bonham-Carter, Geological Survey of Canada, and two anonymous reviewers for helpful suggestions to improve the manuscript.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

REFERENCES

- Agterberg, F. P., and Bonham-Carter, G. F., 2005, Measuring the performance of mineral potential maps: *Nat. Resour. Res.*, v. 14, no. 1, p. 1–17. doi:[10.1007/s11053-005-4674-0](https://doi.org/10.1007/s11053-005-4674-0).
- Arthur, J. D., Wood, H. A. R., Baker, A. F., Cichon, J. R., and Raines, G. L., 2007, Development and implementation of a Bayesian-based aquifer vulnerability assessment in Florida: *Nat. Resour. Res.*, v. 10, no. 2, p. 93–107. doi:[10.1007/s11053-007-9038-5](https://doi.org/10.1007/s11053-007-9038-5).
- Begueria, S., 2006, Validation and evaluation of predictive models in hazard assessment and risk management: *Nat. Hazards*, v. 37, p. 315–329. doi:[10.1007/s11069-005-5182-6](https://doi.org/10.1007/s11069-005-5182-6).
- Behnia, P., 2007, Application of Radial Basis Functional Link Networks to exploration for Proterozoic mineral deposits in central Iran: *Nat. Resour. Res.*, v. 10, no. 2, p. 147–158. doi:[10.1007/s11053-007-9036-7](https://doi.org/10.1007/s11053-007-9036-7).
- Brown, W., Groves, D., and Gedeon, T., 2003, Use of fuzzy membership input layers to combine subjective geological knowledge and empirical data in a neural network method for mineral-potential mapping: *Nat. Resour. Res.*, v. 12, no. 3, p. 183–200. doi:[10.1023/A:1025175904545](https://doi.org/10.1023/A:1025175904545).
- Cheng, Q., 2004, Application of weights of evidence methods for assessment of flowing wells in the Greater Toronto Area, Canada: *Nat. Resour. Res.*, v. 13, no. 2, p. 77–86. doi:[10.1023/B:NARR.0000032645.46747.48](https://doi.org/10.1023/B:NARR.0000032645.46747.48).
- Chung, C. F., 2003, Use of airborne geophysical surveys for constructing mineral potential maps: *Econ. Geol. Monogr.*, v. 11, p. 879–891.
- Chung, C. F., 2006, Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment: *Comput. Geosci.*, v. 32, no. 8, p. 1052–1068. doi:[10.1016/j.cageo.2006.02.003](https://doi.org/10.1016/j.cageo.2006.02.003).
- Chung, C. F., and Fabbri, A. G., 1993, Representation of geoscience data for information integration: *J. Non-Renew. Res.*, v. 2, no. 2, p. 122–139.
- Chung, C. F., and Fabbri, A. G., 1998, Three Bayesian prediction models for landslide hazard, in Buccianti, A., ed., *Proceedings of International Association for Mathematical Geology 1998 Annual Meeting (IAMG'98)*: Ischia, Italy, October 3–7, 1998, p. 204–211.
- Chung, C. F., and Fabbri, A. G., 1999, Probabilistic prediction models for landslide hazard mapping: *Photogramm. Eng. Remote Sens.*, v. 65, no. 12, p. 1389–1399.
- Chung, C. F., and Fabbri, A. G., 2003, Validation of spatial prediction models for landslide hazard mapping, in Chacon, J., and Corominas, J., eds., *Special issue on Landslides and GIS: Nat. Hazards*, v. 30, no. 3, p. 451–472.
- Chung, C. F., and Fabbri, A. G., 2004, Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models, in Glade, T., Anderson, M. G., and Crozier, M. J., eds., *Landslide Hazard and Risk*: John Wiley & Sons, New York, p. 139–174.
- Chung, C. F., and Fabbri, A. G., 2005, On mineral potential maps and how to make them useful, in Cheng, Q., and Bonham-Carter, G., eds., *GIS and Spatial Analysis, Proceedings of IAMG'05*, v. 1: Toronto, Canada, Aug. 21–26, 2005, p. 533–538.
- Chung, C. F., and Fabbri, A. G., 2008, Predicting future landslides for risk analysis—spatial models and cross-validation of their results: *Geomorphology*, v. 94, no. 3–4, p. 438–452. doi:[10.1016/j.geomorph.2006.12.036](https://doi.org/10.1016/j.geomorph.2006.12.036).
- Chung, C. F., Fabbri, A. G., and van Westen, C. J., 1995, Multivariate regression analysis for landslide hazard zonation, in Carrara, A., and Guzzetti, F., eds., *Geographical Information Systems in Assessing Natural Hazards*: Kluwer Academic Publishers, Dordrecht, p. 107–133.
- Chung, C. F., Fabbri, A. G., and Chi, K. H., 2002a, A strategy for sustainable development of nonrenewable resources using spatial prediction models, in Fabbri, A. G., Gáal, G., and McCammon, R. B., eds., *Geoenvironmental Deposit Models for Resource Exploitation and Environmental Security*: Kluwer, Dordrecht, p. 101–118.
- Chung, C. F., Kojima, H., and Fabbri, A. G., 2002, Stability analysis of prediction models for landslide hazard mapping, in Allison, R. J., ed., *Applied Geomorphology: Theory and Practice*: John Wiley and Sons, Ltd., New York, p. 3–19.
- Chung, C. F., Fabbri, A. G., Jang, D. H., and Scholten, H. J., 2005a, Risk assessment using spatial prediction model for natural disaster preparedness, in van Oosterom, P., Zlatanova, S., Fendel, E. M., eds., *Geo-information for Disaster Management*: Springer, Berlin, p. 619–640. *Proceedings of Gi4DM, The First Symposium on Geo-information for Disaster Management*: Delft, Netherlands, March 21–23, 2005, p. 619–640.
- Chung, C. F., Harris, J. R., Keating, P., Kjarsgaard, B., Parsons, S., 2005b, Preliminary report on developing a mineral potential map for diamond exploration in Lac de Gras area, NWT. Unpublished manuscript.
- Chung, C. J., Bonin, D., Fannin, R. J., Fabbri, A. G., Journeay, M., 2006, Uncertainty in predicting landslide hazard in the Coquitlam reservoir area, B.C., Canada, in *Proceedings of IAMG'06*, Liege, Belgium, September 3–8, 2006, Session S09–07, 4 p., CD.
- Coolbaugh, M. F., Raines, G. L., and Zehner, E., 2007, Assessment of exploration bias in data-driven predictive models and the estimation of undiscovered resources: *Nat. Resour. Res.*, v. 10, no. 2, p. 199–207. doi:[10.1007/s11053-007-9037-6](https://doi.org/10.1007/s11053-007-9037-6).
- Corominas, J., Copons, R., Vilaplana, J. M., Altimir, J., and Amigó, J., 2003, Integrated landslide susceptibility analysis and hazard assessment in the principality of Andorra, in Chacon, J., and Corominas, J., eds., *Special issue on Landslides and GIS: Nat. Hazards*, v. 30, no. 3, p. 421–435.
- de Quadros, T. F. P., Koppe, J. C., Strieder, A. J., and Coste, J. F. C. L., 2006, Mineral potential mapping: a comparison of weights-of-evidence and fuzzy methods: *Nat. Resour. Res.*, v. 15, no. 1, p. 49–65. doi:[10.1007/s11053-006-9010-9](https://doi.org/10.1007/s11053-006-9010-9).
- Fabbri, A. G., Chung, C. F., Cendrero, A., and Remondo, J., 2003, Is prediction of future landslides possible with a GIS? in Chacon, J., and Corominas, J., eds., *Special issue on Landslides and GIS: Nat. Hazards*, v. 30, no. 3, p. 487–499.
- Fabbri, A. G., Chung, C. F., and Jang, D. H., 2004, A software approach to spatial predictions of natural hazards and consequent risks, in Brebbia, C. A., ed., *Risk Analysis IV*: WIT Press, Southampton, Boston, p. 289–305.

- Harris, D. V., Zurcher, L., Stanley, M., Marlowe, J., and Pan, G., 2003, A comparative analysis of favorability mapping by weights of evidence, probabilistic neural networks, discriminant analysis and logistic regression: *Nat. Resour. Res.*, v. 12, no. 4, p. 241–255. doi:[10.1023/B:NARR.0000007804.27450.e8](https://doi.org/10.1023/B:NARR.0000007804.27450.e8).
- Lee, S., Ryu, J. H., Lee, M. J., and Won, J. S., 2006, The application of artificial neural networks for landslide susceptibility mapping at Jangshung, Korea: *Math. Geol.*, v. 38, no. 2, p. 199–220. doi:[10.1007/s11004-005-9012-x](https://doi.org/10.1007/s11004-005-9012-x).
- Masetti, M., Poli, S., and Sterlacchini, S., 2007, The use of Weights-of-Evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination: *Nat. Resour. Res.*, v. 10, no. 2, p. 109–119. doi:[10.1007/s11053-007-9045-6](https://doi.org/10.1007/s11053-007-9045-6).
- Nelson, E. P., Connors, K. A., and Suárez, C. S., 2007, GIS-based slope stability analysis, Chuquicamata open pit copper mine, Chile: *Nat. Resour. Res.*, v. 10, no. 2, p. 171–190. doi:[10.1007/s11053-007-9044-7](https://doi.org/10.1007/s11053-007-9044-7).
- Nykanen, V., and Ojala, V. J., 2007, Spatial analysis techniques as successful mineral-potential mapping tools for orogenic gold deposits in the northern Fennoscandian Shield, Finland: *Nat. Resour. Res.*, v. 10, no. 2, p. 85–92. doi:[10.1007/s11053-007-9046-5](https://doi.org/10.1007/s11053-007-9046-5).
- Poli, S., and Sterlacchini, S., 2007, Landslide representation strategies in susceptibility studies using Weights-of-Evidence modeling technique: *Nat. Resour. Res.*, v. 10, no. 2, p. 121–134. doi:[10.1007/s11053-007-9043-8](https://doi.org/10.1007/s11053-007-9043-8).
- Porwal, A., Carranza, E. J. M., and Hale, M., 2003, Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping: *Nat. Resour. Res.*, v. 12, no. 1, p. 1–25. doi:[10.1023/A:1022693220894](https://doi.org/10.1023/A:1022693220894).
- Porwal, A., Carranza, E. J. M., and Hale, M., 2003, Artificial neural networks for mineral-potential mapping: a case study from Aravalli Province, western India: *Nat. Resour. Res.*, v. 12, no. n. 3, p. 155–171. doi:[10.1023/A:1025171803637](https://doi.org/10.1023/A:1025171803637).
- Porwal, A., Carranza, E. J. M., and Hale, M., 2006, Bayesian network classifiers for mineral potential mapping: *Comput. Geosci.*, v. 32, p. 1–16. doi:[10.1016/j.cageo.2005.03.018](https://doi.org/10.1016/j.cageo.2005.03.018).
- Porwal, A., Carranza, E. J. M., and Hale, M., 2006, A hybrid fuzzy weights-of-evidence model for mineral potential mapping: *Nat. Resour. Res.*, v. 15, no. 1, p. 1–14. doi:[10.1007/s11053-006-9012-7](https://doi.org/10.1007/s11053-006-9012-7).
- Raines, G. L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the great basin of the western United States: *Nat. Resour. Res.*, v. 8, no. 4, p. 257–276. doi:[10.1023/A:1021602316101](https://doi.org/10.1023/A:1021602316101).
- Remondo, J., González-Díez, A., Díaz de Terán, J. R., and Cendrero, A., 2003a, Quantitative landslide susceptibility models by means of spatial data analysis techniques; a case study in the lower Deva valley, Guipuzcoa (Spain), *in* Chacon, J., Corominas, J., eds., Special issue on Landslides and GIS: *Nat. Hazards*, v. 30, no. 3, p. 267–279.
- Remondo, J., González-Díez, A., Díaz de Terán, J. R., Cendrero, A., Fabbri, A. G., and Chung, C. F., 2003b, Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain, *in* Chacon, J., Corominas, J., eds., Special issue on Landslides and GIS: *Nat. Hazards*, v. 30, no. 3, p. 437–449.
- Robinson Jr., G. R., and Larkins, P. M., 2007, Probabilistic prediction models for aggregate quarry siting: *Nat. Resour. Res.*, v. 10, no. 2, p. 135–146. doi:[10.1007/s11053-007-9039-4](https://doi.org/10.1007/s11053-007-9039-4).
- Santacana, N., Baeza, C., Corominas, J., de Paz, A., Marturiá, J., 2003, A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet Area (Eastern Pyrenees, Spain), *in* Chacon, J., Corominas, J., eds., Special issue on Landslides and GIS: *Nat. Hazards*, v. 30, no. 3, p. 281–295.
- Singer, D. A., and Kouda, R., 1999, A comparison of weight-of-evidence method and probabilistic neural networks: *Nat. Resour. Res.*, v. 8, no. 4, p. 287–298. doi:[10.1023/A:1021606417010](https://doi.org/10.1023/A:1021606417010).
- Skabar, A. A., 2005, Mapping mineralization probabilities using multilayer perceptrons: *Nat. Resour. Res.*, v. 14, no. 2, p. 109–123. doi:[10.1007/s11053-005-6955-z](https://doi.org/10.1007/s11053-005-6955-z).
- Tissari, S., Nykänene, V., Lerssi, J., and Kolehmainen, M., 2007, Classification of soil groups using Weights-of-Evidence method and RBFLN-neural nets: *Nat. Resour. Res.*, v. 10, no. 2, p. 159–169. doi:[10.1007/s11053-007-9040-y](https://doi.org/10.1007/s11053-007-9040-y).
- van Westen, C. J., Rengers, N., Soeters, R., 2003, Use of geomorphological information in indirect landslide susceptibility assessment. *in* Chacon, J., Corominas, J., eds., Special issue on Landslides and GIS: *Nat. Hazards*, v. 30, no. 3, p. 399–419.
- Zêzere, J. L., Reis, E., Garcia, R., Oliveira, S., Rodrigues, M. L., Vieira, G., and Ferreira, A. B., 2004, Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal): *Nat. Hazards Earth Syst. Sci.*, v. 4, p. 133–146.